PORTAL USPTO

Subscribe (Full Service) Register (Limited Service, Free) Login

Search: The ACM Digital Library The Guide

Automated Text Summarization and term vector frequency

THE ACM DIGITAL LIBRARY

Feedback Report a problem Satisfaction survey

Terms used
Automated Text Summarization and **term vector frequency**

Found **38,902** of **169,166**

Sort results by: relevance
Display results: expanded form

Save results to a Binder
Search Tips
Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 20 of 200 Result page: **1** 2 3 4 5 6 7 8 9 10 next
Best 200 shown

Relevance scale

1 Recognising and using named entities: Focused named entity recognition using machine learning
Li Zhang, Yue Pan, Tong Zhang
July 2004 **Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR '04**
Publisher: ACM Press
Full text available: pdf(208.61 KB) Additional Information: full citation, abstract, references, index terms

In this paper we study the problem of finding most topical named entities among all entities in a document, which we refer to as focused named entity recognition. We show that these focused named entities are useful for many natural language processing applications, such as document summarization, search result ranking, and entity detection and tracking. We propose a statistical model for focused named entity recognition by converting it into a classification problem. We then study the impact of ...

Keywords: decision tree, information retrieval, naive bayes, robust risk minimization, text summarization, topic identification

2 A document retrieval model based on term frequency ranks
Ijsbrand Jan Aalbersberg
August 1994 **Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval**
Publisher: Springer-Verlag New York, Inc.
Full text available: pdf(951.78 KB) Additional Information: full citation, references, citings, index terms, review

3 Novel search environments: Comparison of two approaches to building a vertical search tool: a case study in the nanotechnology domain
Michael Chau, Hsinchun Chen, Jialun Qin, Yilu Zhou, Yi Qin, Wai-Ki Sung, Daniel McDonald
July 2002 **Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries**
Publisher: ACM Press
Full text available: pdf(859.29 KB) Additional Information: full citation, abstract, references, citings, index terms

As the Web has been growing exponentially, it has become increasingly difficult to search for desired information. In recent years, many domain-specific (vertical) search tools have been developed to serve the information needs of specific fields. This paper describes two approaches to building a domain-specific search tool. We report our experience in building

two different tools in the nanotechnology domain -- (1) a server-side search engine, and (2) a client-side search agent. The designs of ...

Keywords: indexing, information retrieval, internet searching and browsing, internet spider, noun-phrasing, personalization, post-retrieval analysis, self-organizing map, summarization, vertical search engine, web search engine

4 Using suffix arrays to compute term frequency and document frequency for all substrings in a corpus

Mikio Yamamoto, Kenneth W. Church
March 2001 **Computational Linguistics**, Volume 27 Issue 1
Publisher: MIT Press
Full text available: pdf(1.80 MB) Additional Information: full citation, abstract, references, citings
Publisher Site

Bigrams and trigrams are commonly used in statistical natural language processing; this paper will describe techniques for working with much longer *n*-grams. Suffix arrays (Manber and Myers 1990) were first introduced to compute the frequency and location of a substring (*n*-gram) in a sequence (corpus) of length *N*. To compute frequencies over all *N(N + 1)/2* substrings in a corpus, the substrings are grouped into a manageable number of equivalence classes. In this way, a ...

5 Set-based vector model: An efficient approach for correlation-based ranking

Bruno Pôssas, Nivio Ziviani, Wagner Meira, Berthier Ribeiro-Neto
October 2005 **ACM Transactions on Information Systems (TOIS)**, Volume 23 Issue 4
Publisher: ACM Press
Full text available: pdf(800.89 KB) Additional Information: full citation, abstract, references, index terms

This work presents a new approach for ranking documents in the vector space model. The novelty lies in two fronts. First, patterns of term co-occurrence are taken into account and are processed efficiently. Second, term weights are generated using a data mining technique called association rules. This leads to a new ranking mechanism called the *set-based vector model*. The components of our model are no longer index terms but index termsets, where a termset is a set of index terms. Termset ...

Keywords: Information retrieval models, association rule mining, correlation-based ranking, data mining, weighting index term co-occurrences

6 Meaningful term extraction and discriminative term selection in text categorization via unknown-word methodology

Yu-Sheng Lai, Chung-Hsien Wu
March 2002 **ACM Transactions on Asian Language Information Processing (TALIP)**, Volume 1 Issue 1
Publisher: ACM Press
Full text available: pdf(920.43 KB) Additional Information: full citation, abstract, references, index terms

In this article, an approach based on unknown words is proposed for meaningful term extraction and discriminative term selection in text categorization. For meaningful term extraction, a phrase-like unit (PLU)-based likelihood ratio is proposed to estimate the likelihood that a word sequence is an unknown word. On the other hand, a discriminative measure is proposed for term selection and is combined with the PLU-based likelihood ratio to determine the text category. We conducted several experim ...

Keywords: AC-machine, dimensionality reduction, discriminability, discriminative term selection, inconsistency problem, meaningful term extraction, n-gram, phrase-like unit, sparse data problem, term adaptation, term purification, text categorization, text indexing, unknown word detection, vector space modeling

7 Information retrieval session 1: adhoc retrieval: A study of parameter tuning for term frequency normalization

Ben HE, Iadh Ounis

November 2003 **Proceedings of the twelfth international conference on Information and knowledge management**

Publisher: ACM Press

Full text available: pdf(152.19 KB) Additional Information: full citation, abstract, references, citings, index terms

Most current term frequency normalization approaches for information retrieval involve the use of parameters. The tuning of these parameters has an important impact on the overall performance of the information retrieval system. Indeed, a small variation in the involved parameter(s) could lead to an important variation in the precision/recall values. Most current tuning approaches are dependent on the document collections. As a consequence, the effective parameter value cannot be obtained for a ...

Keywords: document length, information retrieval, parameter tuning, term frequency normalization

8 A vector space model for automatic indexing

G. Salton, A. Wong, C. S. Yang

November 1975 **Communications of the ACM**, Volume 18 Issue 11

Publisher: ACM Press

Full text available: pdf(687.42 KB) Additional Information: full citation, abstract, references, citings, index terms

In a document retrieval, or other pattern matching environment where stored entities (documents) are compared with each other or with incoming patterns (search requests), it appears that the best indexing (property) space is one where each entity lies as far away from the others as possible; in these circumstances the value of an indexing system may be expressible as a function of the density of the object space; in particular, retrieval performance may correlate inversely with space densit ...

Keywords: automatic indexing, automatic information retrieval, content analysis, document space

9 Effective information retrieval using term accuracy

C. T. Yu, G. Salton

March 1977 **Communications of the ACM**, Volume 20 Issue 3

Publisher: ACM Press

Full text available: pdf(799.09 KB) Additional Information: full citation, abstract, references, citings

The performance of information retrieval systems can be evaluated in a number of different ways. Much of the published evaluation work is based on measuring the retrieval performance of an average user query. Unfortunately, formal proofs are difficult to construct for the average case. In the present study, retrieval evaluation is based on optimizing the performance of a specific user query. The concept of query term accuracy is introduced as the probability of occurrence of a query term in ...

Keywords: automatic indexing, content analysis, frequency weighting, information retrieval, term accuracy, thesaurus and phrase transformations

10 A statistical learning learning model of text classification for support vector machines

Thorsten Joachims

September 2001 **Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(293.19 KB) Additional Information: full citation, abstract, references, citings, index terms

This paper develops a theoretical learning model of text classification for Support Vector Machines (SVMs). It connects the statistical properties of text-classification tasks with the generalization performance of a SVM in a quantitative way. Unlike conventional approaches to learning text classifiers, which rely primarily on empirical evidence, this model explains why and when SVMs perform well for text classification. In particular, it addresses the following questions: Why can support v ...

11 Vector-based natural language call routing

Jennifer Chu-Carroll, Bob Carpenter

September 1999 **Computational Linguistics**, Volume 25 Issue 3

Publisher: MIT Press

Full text available: pdf(1.87 MB) Publisher Site Additional Information: full citation, abstract, references, citings

This paper describes a domain-independent, automatically trained natural language call router for directing incoming calls in a call center. Our call router directs customer calls based on their response to an open-ended *How may I direct your call?* prompt. Routing behavior is trained from a corpus of transcribed and hand-routed calls and then carried out using vector-based information retrieval techniques. Terms consist of *n*-gram sequences of morphologically reduced content words, ...

12 Frequency space environment map rendering

Ravi Ramamoorthi, Pat Hanrahan

July 2002 **ACM Transactions on Graphics (TOG) , Proceedings of the 29th annual conference on Computer graphics and interactive techniques SIGGRAPH '02**, Volume 21 Issue 3

Publisher: ACM Press

Full text available: pdf(3.37 MB) Additional Information: full citation, abstract, references, citings, index terms

We present a new method for real-time rendering of objects with complex isotropic BRDFs under distant natural illumination, as specified by an environment map. Our approach is based on spherical frequency space analysis and includes three main contributions. Firstly, we are able to theoretically analyze required sampling rates and resolutions, which have traditionally been determined in an ad-hoc manner. We also introduce a new compact representation, which we call a *spherical harmonic reflec ...*

Keywords: *complexity analysis, environment maps, image-based rendering, signal-processing, spherical harmonics*

13 Evaluation of the 2-Poisson model as a basis for using term frequency data in searching

Vijay V. Raghavan, Hong-pao Shi, C. T. Yu

June 1983 **ACM SIGIR Forum , Proceedings of the 6th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR '83**, Volume 17 Issue 4

Publisher: ACM Press

Full text available: pdf(971.22 KB) Additional Information: full citation, abstract, references, citings

The early work on the probabilistic models of retrieval assumed that the document representation is binary, indicating only the presence or absence of index terms. The 2-Poisson (TP) model which was proposed as a model of how the occurrence frequency of specialty words in a collection is distributed, has since been used to develop retrieval strategies that incorporate term frequency information. This work investigates the use of the TP model, in this context, further. It is shown that the search ...

14 On the update of term weights in dynamic information retrieval systems

Charles L. Viles, James C. French

December 1995 **Proceedings of the fourth international conference on Information and knowledge management**

Publisher: ACM Press

Full text available: pdf(834.31 KB) Additional Information: full citation, references, citings, index terms

15 On modeling of information retrieval concepts in vector spaces

S. K.M. Wong, W. Ziarko, V. V. Raghavan, P. C.N. Wong

June 1987 **ACM Transactions on Database Systems (TODS)**, Volume 12 Issue 2

Publisher: ACM Press

Full text available: pdf(1.80 MB) Additional Information: full citation, abstract, references, citings, index terms, review

The Vector Space Model (VSM) has been adopted in information retrieval as a means of coping with inexact representation of documents and queries, and the resulting difficulties in determining the relevance of a document relative to a given query. The major problem in employing this approach is that the explicit representation of term vectors is not known a priori. Consequently, earlier researchers made the assumption that the vectors corresponding to terms are pairwise orthogonal. Such an a ...

16 Poster papers: CVS: a Correlation-Verification based Smoothing technique on information retrieval and term clustering

Christina Yip Chung, Bin Chen

July 2002 **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining**

Publisher: ACM Press

Full text available: pdf(634.38 KB) Additional Information: full citation, abstract, references, index terms

As information volume in enterprise systems and in the Web grows rapidly, how to accurately retrieve information is an important research area. Several corpus based smoothing techniques have been proposed to address the data sparsity and synonym problems faced by information retrieval systems. Such smoothing techniques are often unable to discover and utilize the correlations among terms.We propose CVS, a Correlation-Verification based Smoothing method, that considers co-occurrence information i ...

Keywords: information retrieval, query expansion, smoothing, term clustering, text mining

17 Vector space model of information retrieval: a reevaluation

S. K. M. Wong, Vijay V. Raghavan

July 1984 **Proceedings of the 7th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: British Computer Society

Full text available: pdf(832.89 KB) Additional Information: full citation, abstract, references, citings

In this paper we, in essence, point out that the methods used in the current vector based systems are in conflict with the premises of the vector space model. The considerations, naturally, lead to how things might have been done differently. More importantly, it is felt that this investigation will lead to a clearer understanding of the issues and problems in using the vector space model in information retrieval.

18 Orthogonal negation in vector spaces for modelling word-meanings and document retrieval

Dominic Widdows

July 2003 **Proceedings of the 41st Annual Meeting on Association for Computational**

Linguistics - Volume 1 ACL '03
Publisher: Association for Computational Linguistics
Full text available: pdf(168.48 KB) Additional Information: full citation, abstract, references

Standard IR systems can process queries such as "web NOT internet", enabling users who are interested in arachnids to avoid documents about computing. The documents retrieved for such a query should be irrelevant to the negated query term. Most systems implement this by reprocessing results after retrieval to remove documents containing the unwanted string of letters.This paper describes and evaluates a theoretically motivated method for removing unwanted meanings directly from the original quer ...

19 A comparison of search term weighting: term relevance vs. inverse document frequency
Harry Wu, Gerard Salton
May 1981 **ACM SIGIR Forum , Proceedings of the 4th annual international ACM SIGIR conference on Information storage and retrieval: theoretical issues in information retrieval SIGIR '81**, Volume 16 Issue 1
Publisher: ACM Press
Full text available: pdf(563.09 KB) Additional Information: full citation, abstract, references, citings

The term relevance weighting method has been shown to produce optimal information retrieval queries under well-defined conditions. The parameters needed to generate the term relevance factors cannot unfortunately be estimated accurately in practice; futhermore, in realistic test situations, it appears difficult to obtain improved retrieval results using the term relevance weights over much simpler term weighting systems such as, for example, the inverse document frequency weights.It is shown in ...

20 All-frequency shadows using non-linear wavelet lighting approximation
Ren Ng, Ravi Ramamoorthi, Pat Hanrahan
July 2003 **ACM Transactions on Graphics (TOG)**, Volume 22 Issue 3
Publisher: ACM Press
Full text available: pdf(5.22 MB) Additional Information: full citation, abstract, references, citings, index terms

We present a method, based on pre-computed light transport, for real-time rendering of objects under all-frequency, time-varying illumination represented as a high-resolution environment map. Current techniques are limited to small area lights, with sharp shadows, or large low-frequency lights, with very soft shadows. Our main contribution is to approximate the environment map in a wavelet basis, keeping only the largest terms (this is known as a *non-linear approximation*). We obtain furth ...

Keywords: image-based rendering, non-linear approximation, relighting, shadow algorithms, spherical harmonics, wavelets

Results 1 - 20 of 200 Result page: **1** 2 3 4 5 6 7 8 9 10 next

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2006 ACM, Inc.
Terms of Usage Privacy Policy Code of Ethics Contact Us

Useful downloads: Adobe Acrobat QuickTime Windows Media Player Real Player

PORTAL USPTO
Subscribe (Full Service) Register (Limited Service, Free) Login
Search: The ACM Digital Library The Guide
sentences and term vector frequency
THE ACM DIGITAL LIBRARY
Feedback Report a problem Satisfaction survey

Terms used **sentences** and **term vector frequency** Found **44,965** of **169,166**

Sort results by: relevance
Display results: expanded form
Save results to a Binder
Search Tips
Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 121 - 140 of 200 Result page: previous 1 2 3 4 5 6 **7** 8 9 10 next
Best 200 shown Relevance scale

121 A study of probability kinematics in information retrieval
F. Crestani, C. J. van Rijsbergen
July 1998 **ACM Transactions on Information Systems (TOIS)**, Volume 16 Issue 3
Publisher: ACM Press
Full text available: pdf(1.12 MB) Additional Information: full citation, abstract, references, citings, index terms, review

We analyze the kinematics of probabilistic term weights at retrieval time for different Information Retrieval models. We present four models based on different notions of probabilistic retrieval. Two of these models are based on classical probability theory and can be considered as prototypes of models long in use in Information Retrieval, like the Vector Space Model and the Probabilistic Model. The two other models are based on a logical technique of evaluating the probability of a conditi ...

Keywords: logical imaging, probabilistic modeling, probabilistic retrieval

122 Automated techniques for managing collections: Managing distributed collections: evaluating web page changes, movement, and replacement
Zubin Dalal, Suvendu Dash, Pratik Dave, Luis Francisco-Revilla, Richard Furuta, Unmil Karadkar, Frank Shipman
June 2004 **Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries**
Publisher: ACM Press
Full text available: pdf(329.43 KB) Additional Information: full citation, abstract, references, citings, index terms

Distributed collections of Web materials are common. Bookmark lists, paths, and catalogs such as Yahoo! Directories require human maintenance to keep up to date with changes to the underlying documents. The Walden's Paths Path Manager is a tool to support the maintenance of distributed collections. Earlier efforts focused on recognizing the type and degree of change within Web pages and identifying pages no longer accessible. We now extend this work with algorithms for evaluating drastic changes ...

Keywords: change detection, collection management, document location

123 Discovering all most specific sentences
Dimitrios Gunopulos, Roni Khardon, Heikki Mannila, Sanjeev Saluja, Hannu Toivonen, Ram Sewak Sharma
June 2003 **ACM Transactions on Database Systems (TODS)**, Volume 28 Issue 2
Publisher: ACM Press
Additional Information:

Full text available: pdf(283.09 KB) full citation, abstract, references, citings, index terms

Data mining can be viewed, in many instances, as the task of computing a representation of a theory of a model or a database, in particular by finding a set of maximally specific sentences satisfying some property. We prove some hardness results that rule out simple approaches to solving the problem.The *a priori* algorithm is an algorithm that has been successfully applied to many instances of the problem. We analyze this algorithm, and prove that is optimal when the maximally specific sen ...

Keywords: Data mining, association rules, learning with membership queries, maximal frequent sets, minimal keys

124 A system for retrieving speech documents

Ulrike Glavitsch, Peter Schäuble

June 1992 **Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(941.62 KB) Additional Information: full citation, abstract, references, citings, index terms

An information retrieval model is presented for the retrieval of speech documents, i.e. audio recordings containing speech. The indexing vocabulary consists of indexing features that have the following characteristics. First, they are easy to recognize by speech recognition methods. Second, the number of different indexing features is small such that a reasonable amount of training data is sufficent to train the hidden Markov models that are used by the speech recognition process. Third, th ...

125 Lexical ambiguity and information retrieval

Robert Krovetz, W. Bruce Croft

April 1992 **ACM Transactions on Information Systems (TOIS)**, Volume 10 Issue 2

Publisher: ACM Press

Full text available: pdf(2.00 MB) Additional Information: full citation, abstract, references, citings, index terms, review

Lexical ambiguity is a pervasive problem in natural language processing. However, little quantitative information is available about the extent of the problem or about the impact that it has on information retrieval systems. We report on an analysis of lexical ambiguity in information retrieval test collections and on experiments to determine the utility of word meanings for separating relevant from nonrelevant documents. The experiments show that there is considerable ambiguity even in a s ...

Keywords: disambiguation, document retrieval, semantically based search, word senses

126 Special issue on independent components analysis: Energy-based models for sparse overcomplete representations

Yee Whye Teh, Max Welling, Simon Osindero, Geoffrey E. Hinton

December 2003 **The Journal of Machine Learning Research**, Volume 4

Publisher: MIT Press

Full text available: pdf(591.75 KB) Additional Information: full citation, abstract, index terms

We present a new way of extending independent components analysis (ICA) to overcomplete representations. In contrast to the causal generative extensions of ICA which maintain marginal independence of *sources*, we define *features* as deterministic (linear) functions of the inputs. This assumption results in marginal *dependencies* among the features, but *conditional independence* of the features given the inputs. By assigning energies to the features a probability d ...

127 Cosmological lower bound on the circuit complexity of a small problem in logic

Larry Stockmeyer, Albert R. Meyer
November 2002 **Journal of the ACM (JACM)**, Volume 49 Issue 6
Publisher: ACM Press
Full text available: pdf(277.59 KB) Additional Information: full citation, abstract, references, index terms

An exponential lower bound on the circuit complexity of deciding the weak monadic second-order theory of one successor (WS1S) is proved. Circuits are built from binary operations, or 2-input gates, which compute arbitrary Boolean functions. In particular, to decide the truth of logical formulas of length at most 610 in this second-order language requires a circuit containing at least 10^{125} gates. So even if each gate were the size of a proton, the circuit would not fit in the known un ...

Keywords: Circuit complexity, WS1S, computational complexity, decision problem, logic, lower bound, practical undecidability

128 Human memory models and term association

Gerda Ruge
July 1995 **Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval**
Publisher: ACM Press
Full text available: pdf(936.33 KB) Additional Information: full citation, references, citings, index terms

129 A document retrieval system for man-machine interaction

G. Salton
January 1964 **Proceedings of the 1964 19th ACM national conference**
Publisher: ACM Press
Full text available: pdf(1.07 MB) Additional Information: full citation, abstract, references, citings, index terms

An automatic document retrieval system, programmed for the IBM 7094, is described. The system is designed to process English texts and search requests, and uses statistical, syntactic and semantic procedures for the analysis of information and the identification of relevant items. The operations are planned around a central supervisor, which in turn calls on the various subroutines, as desired. This organization makes it possible to alter both the processing sequences and the mat ...

130 Mood swings: expressive speech animation

Erika Chuang, Christoph Bregler
April 2005 **ACM Transactions on Graphics (TOG)**, Volume 24 Issue 2
Publisher: ACM Press
Full text available: pdf(8.58 MB) Additional Information: full citation, abstract, references, index terms

Motion capture-based facial animation has recently gained popularity in many applications, such as movies, video games, and human-computer interface designs. With the use of sophisticated facial motions from a human performer, animated characters are far more lively and convincing. However, editing motion data is difficult, limiting the potential of reusing the motion data for different tasks. To address this problem, statistical techniques have been applied to learn models of the facial motion ...

Keywords: Facial animation, expression, motion, retargeting

131 Information retrieval models: Inferring query models by computing information flow

P. D. Bruza, D. Song
November 2002 **Proceedings of the eleventh international conference on Information and knowledge management**
Publisher: ACM Press

Full text available: pdf(278.90 KB) Additional Information: full citation, abstract, references, citings, index terms

The language modelling approach to information retrieval can also be used to compute query models. A query model can be envisaged as an expansion of an initial query. The more prominent query models in the literature have a probabilistic basis. This paper introduces an alternative, non-probabilistic approach to query modelling whereby the strength of information flow is computed between a query *Q* and a term *w*. Information flow is a reflection of how strongly *w* is *informat ...*

***Keywords**: inference, information flow, query language modelling*

132 Special issue on using large corpora: I: Generalized probabilistic LR parsing of natural language (Corpora) with unification-based grammars

Ted Briscoe, John Carroll

March 1993 **Computational Linguistics**, Volume 19 Issue 1

Publisher: MIT Press

Full text available: pdf(2.62 MB) Additional Information: full citation, abstract, references, citings

Publisher Site

We describe work toward the construction of a very wide-coverage probabilistic parsing system for natural language (NL), based on LR parsing techniques. The system is intended to rank the large number of syntactic analyses produced by NL grammars according to the frequency of occurrence of the individual rules deployed in each analysis. We discuss a fully automatic procedure for constructing an LR parse table from a unification-based grammar formalism, and consider the suitability of alternative ...

133 Automatic corpus-based tone and break-index prediction using K-ToBI representation

Jin-Seok Lee, Byeongchang Kim, Gary Geunbae Lee

September 2002 **ACM Transactions on Asian Language Information Processing (TALIP)**, Volume 1 Issue 3

Publisher: ACM Press

Full text available: pdf(149.55 KB) Additional Information: full citation, abstract, references, index terms

In this article we present a prosody generation architecture based on K-ToBI (Korean Tone and Break Index) representation. ToBI is a multitier representation system based on linguistic knowledge that transcribes events in an utterance. The TTS (Text-To-Speech) system, which adopts ToBI as an intermediate representation, is known to exhibit higher flexibility, modularity, and domain/task portability compared to the direct prosody generation TTS systems. However, for practical-level performance, t ...

Keywords: K-ToBI, intonation, phrase break, pitch, prosodic phrase, prosody, text-to-speech system

134 Paper session IR-6 (information retrieval): IR models 1: Word sense disambiguation in queries

Shuang Liu, Clement Yu, Weiyi Meng

October 2005 **Proceedings of the 14th ACM international conference on Information and knowledge management CIKM '05**

Publisher: ACM Press

Full text available: pdf(143.44 KB) Additional Information: full citation, abstract, references, index terms

This paper presents a new approach to determine the senses of words in queries by using WordNet. In our approach, noun phrases in a query are determined first. For each word in the query, information associated with it, including its synonyms, hyponyms, hypernyms, definitions of its synonyms and hyponyms, and its domains, can be used for word sense disambiguation. By comparing these pieces of information associated with the words which form a phrase, it may be possible to assign senses to these ...

Keywords: WordNet, information retrieval, word sense disambiguation

135 Word sense disambiguation using static and dynamic sense vectors

Jong-Hoon Oh, Key-Sun Choi

August 2002 **Proceedings of the 19th international conference on Computational linguistics - Volume 1**

Publisher: Association for Computational Linguistics

Full text available: pdf(98.87 KB) Additional Information: full citation, abstract, references

It is popular in WSD to use contextual information in training sense tagged data. Co-occurring words within a limited window-sized context support one sense among the semantically ambiguous ones of the word. This paper reports on word sense disambiguation of English words using static and dynamic sense vectors. First, context vectors are constructed using contextual words in the training sense tagged data. Then, the words in the context vector are weighted with local density. Using the whole tra ...

136 Oral I: Person identification using automatic integration of speech, lip, and face experts

Niall A. Fox, Ralph Gross, Philip de Chazal, Jeffery F. Cohn, Richard B. Reilly

November 2003 **Proceedings of the 2003 ACM SIGMM workshop on Biometrics methods and applications**

Publisher: ACM Press

Full text available: pdf(293.18 KB) Additional Information: full citation, abstract, references, index terms

This paper presents a multi-expert person identification system based on the integration of three separate systems employing audio features, static face images and lip motion features respectively. Audio person identification was carried out using a text dependent Hidden Markov Model methodology. Modeling of the lip motion was carried out using Gaussian probability density functions. The static image based identification was carried out using the FaceIt system. Experiments were conducted with 25 ...

Keywords: audio, automatic weighting, face, late integration, lips, multi-expert, person identification

137 Automatic text indexing using complex identifiers

Gerald Salton

January 2000 **Proceedings of the ACM conference on Document processing systems**

Publisher: ACM Press

Full text available: pdf(569.46 KB) Additional Information: full citation, references, citings, index terms

138 Book reviews: Review of "Parallel text processing: alignment and use of translation corpora" by Jean Véronis. Kluwer Academic Publishers 2000.

Philip Resnik

December 2001 **Computational Linguistics**, Volume 27 Issue 4

Publisher: MIT Press

Full text available: pdf(340.92 KB) Publisher Site Additional Information: full citation, references

139 Computer-aided law and advanced technologies (CLAT): Sequence modelling for sentence classification in a legal summarisation system

Ben Hachey, Claire Grover

March 2005 **Proceedings of the 2005 ACM symposium on Applied computing**

Publisher: ACM Press
Full text available: pdf(105.27 KB) Additional Information: full citation, abstract, references

We describe a set of experiments using a wide range of machine learning techniques for the task of predicting the rhetorical status of sentences. The research is part of a text summarisation project for the legal domain for which we use a new corpus of judgments of the UK House of Lords. We present experimental results for classification according to a rhetorical scheme indicating a sentence's contribution to the overall argumentative structure of the legal judgments using four learning algorith ...

Keywords: artificial intelligence, automatic summarisation, discourse, law, natural language

140 Computer Evaluation of Indexing and Text Processing

G. Salton, M. E. Lesk
January 1968 **Journal of the ACM (JACM)**, Volume 15 Issue 1
Publisher: ACM Press
Full text available: pdf(2.19 MB) Additional Information: full citation, abstract, references, citings, index terms

Automatic indexing methods are evaluated and design criteria for modern information systems are derived.

Results 121 - 140 of 200 Result page: previous 1 2 3 4 5 6 **7** 8 9 10 next

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2006 ACM, Inc.
Terms of Usage Privacy Policy Code of Ethics Contact Us

Useful downloads: Adobe Acrobat QuickTime Windows Media Player Real Player

Terms used **sentences** and **term** **vector** **frequency**

Found **44,965** of **169,166**

Sort results by: relevance

Display results: expanded form

Save results to a Binder

Search Tips

☐ Open results in a new window

Try an Advanced Search

Try this search in The ACM Guide

Results 21 - 40 of 200 Result page: previous 1 **2** 3 4 5 6 7 8 9 10 next

Best 200 shown

Relevance scale

21 Automatic text structuring and retrieval-experiments in automatic encyclopedia searching

Gerard Salton, Chris Buckley

September 1991 **Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(947.93 KB) Additional Information: full citation, references

approaches are described in this study for implemen ...

24 Building efficient and effective metasearch engines

Weiyi Meng, Clement Yu, King-Lup Liu

March 2002 **ACM Computing Surveys (CSUR)**, Volume 34 Issue 1

Publisher: ACM Press

Full text available: pdf(416.07 KB) Additional Information: full citation, abstract, references, citings, index terms

Frequently a user's information needs are stored in the databases of multiple search engines. It is inconvenient and inefficient for an ordinary user to invoke multiple search engines and identify useful documents from the returned results. To support unified access to multiple search engines, a metasearch engine can be constructed. When a metasearch engine receives a query from a user, it invokes the underlying search engines to retrieve useful information for the user. Metasearch engines have ...

Keywords: Collection fusion, distributed collection, distributed information retrieval, information resource discovery, metasearch

25 Knowledge base maintenance using knowledge gap analysis

Scott Spangler, Jeffrey Kreulen

August 2001 **Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining**

Publisher: ACM Press

Full text available: pdf(491.63 KB) Additional Information: full citation, abstract, references, index terms

As the web and e-business have proliferated, the practice of using customer facing knowledge bases to augment customer service and support operations has increased. This can be a very efficient, scalable and cost effective way to share knowledge. The effectiveness and cost savings are proportional to the utility of the information within the knowledge base and inversely proportional to the amount of labor required in maintaining the knowledge. To address this issue, we have developed an algorith ...

Keywords: Clustering, Gap Analysis, Knowledge Management, Text Mining

26 Methodology, performance: Automatic metadata generation based on neural network

Yunfeng Li, Qingsheng Zhu, Yukun Cao

November 2004 **Proceedings of the 3rd international conference on Information security InfoSecu '04**

Publisher: ACM Press

Full text available: pdf(459.72 KB) Additional Information: full citation, abstract, references, index terms

Metadata technologies can allow the proper search and processing of Web pages, which has been used in many fields, such as e-commerce, library science, web search, and so on. In the approach, a novel framework is proposed to automatically generate Dublin Core metadata about web pages, what is finally expressed in XML. The framework is located on a web server, where can provide detailed information about web pages to generate the metadata. Moreover, the framework utilizes a combination of a well- ...

Keywords: PCA, RDF, dublin core, neural network

27 Keyword extraction of radio news using term weighting with an encyclopedia and newspaper articles

Fumiyo Fukumoto, Yoshihiro Sekiguchi

August 1998 **Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(232.78 KB) Additional Information: full citation, references, index terms

28 Automatic phrase indexing for document retrieval

J. Fagan
November 1987 **Proceedings of the 10th annual international ACM SIGIR conference on Research and development in information retrieval**
Publisher: ACM Press

Full text available: pdf(1.10 MB) Additional Information: full citation, abstract, references, citings, index terms

An automatic phrase indexing method based on the term discrimination model is described, and the results of retrieval experiments on five document collections are presented. Problems related to this non-syntactic phrase construction method are discussed, and some possible solutions are proposed that make use of information about the syntactic structure of document and query texts.

29 Technique for automatically correcting words in text

Karen Kukich
December 1992 **ACM Computing Surveys (CSUR)**, Volume 24 Issue 4
Publisher: ACM Press

Full text available: pdf(6.23 MB) Additional Information: full citation, abstract, references, citings, index terms, review

Research aimed at correcting words in text has focused on three progressively more difficult problems:(1) nonword error detection; (2) isolated-word error correction; and (3) context-dependent work correction. In response to the first problem, efficient pattern-matching and n-gram analysis techniques have been developed for detecting strings that do not appear in a given word list. In response to the second problem, a variety of general and application-specific spelling cor ...

Keywords: n-gram analysis, Optical Character Recognition (OCR), context-dependent spelling correction, grammar checking, natural-language-processing models, neural net classifiers, spell checking, spelling error detection, spelling error patterns, statistical-language models, word recognition and correction

30 A corpus analysis approach for automatic query expansion and its extension to multiple databases

Susan Gauch, Jianying Wang, Satya Mahesh Rachakonda
July 1999 **ACM Transactions on Information Systems (TOIS)**, Volume 17 Issue 3
Publisher: ACM Press

Full text available: pdf(111.47 KB) Additional Information: full citation, abstract, references, citings, index terms

Searching online text collections can be both rewarding and frustrating. While valuable information can be found, typically many irrelevant documents are also retrieved, while many relevant ones are missed. Terminology mismatches between the user's query and document contents are a main cause of retrieval failures. Expanding a user's query with related words can improve search performances, but finding and using related words is an open problem. This research uses corpus analysis technique ...

Keywords: query expansion

31 Posters: Summary evaluation and text categorization

Khurshid Ahmad, Bogdan Vrusias, Paulo C F de Oliveira
July 2003 **Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval**
Publisher: ACM Press

Full text available: pdf(177.75 KB) Additional Information: full citation, abstract, references, index terms

In general terms the evaluation of a summary depends on how close it is to the chief points in the source text. This begets the question as to what are the chief points in the source text and how is this information used in itself in identifying the source text. This is crucially important when we discuss automatic evaluation of summaries. So the question of main points is the source text. Typically, this would be around a nucleus of keywords. However, the salience, the frequency, and the relati ...

Keywords: summary evaluation, text categorization, training vectors

32 A Chinese dictionary construction algorithm for information retrieval

Honglan Jin, Kam-Fai Wong

December 2002 **ACM Transactions on Asian Language Information Processing (TALIP)**, Volume 1 Issue 4

Publisher: ACM Press

Full text available: pdf(133.47 KB) Additional Information: full citation, abstract, references, index terms

In this article we propose a method for constructing, from raw Chinese text, a statistics-based automatic dictionary. The method makes use of local statistical information (i.e., data within a document) to identify and discard repeated string patterns, which, at an earlier stage, were substrings of legitimate words. Global statistical information (which exists throughout the entire corpus) and contextual constraints are then used for further filtering. The method can be used to alleviate the out ...

Keywords: Chinese information retrieval, automatic word extraction, dictionary construction

33 Text categorization for multiple users based on semantic features from a machine-readable dictionary

Elizabeth D. Liddy, Woojin Paik, Edmund S. Yu

July 1994 **ACM Transactions on Information Systems (TOIS)**, Volume 12 Issue 3

Publisher: ACM Press

Full text available: pdf(1.17 MB) Additional Information: full citation, abstract, references, citings, index terms, review

The text categorization module described here provides a front-end filtering function for the larger DR-LINK text retrieval system [Liddy and Myaeing 1993]. The model evaluates a large incoming stream of documents to determine which documents are sufficiently similar to a profile at the broad subject level to warrant more refined representation and matching. To accomplish this task, each substantive word in a text is first categorized using a feature set based on the semantic Subject Field ...

Keywords: semantic vectors, subject field coding

34 Automatic combination of multiple ranked retrieval systems

Brian T. Bartell, Garrison W. Cottrell, Richard K. Belew

August 1994 **Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: Springer-Verlag New York, Inc.

Full text available: pdf(834.64 KB) Additional Information: full citation, references, citings, index terms

35 Towards effective strategies for monolingual and bilingual information retrieval: Lessons learned from NTCIR-4

Yan Qu, David A. Hull, Gregory Grefenstette, David A. Evans, Motoko Ishikawa, Setsuko Nara, Toshiya Ueda, Daisuke Noda, Kousaku Arita, Yuki Funakoshi, Hiroshi Matsuda

June 2005 **ACM Transactions on Asian Language Information Processing (TALIP)**, Volume 4 Issue 2
Publisher: ACM Press
Full text available: pdf(502.73 KB) Additional Information: full citation, abstract, references, index terms

At the NTCIR-4 workshop, Justsystem Corporation (JSC) and Clairvoyance Corporation (CC) collaborated in the cross-language retrieval task (CLIR). Our goal was to evaluate the performance and robustness of our recently developed commercial-grade CLIR systems for English and Asian languages. The main contribution of this article is the investigation of different strategies, their interactions in both monolingual and bilingual retrieval tasks, and their respective contributions to operational retri ...

Keywords: Monolingual information retrieval, NTCIR, comparison, cross-language information retrieval

36 Technical papers: design recovery and documentation: Recovering documentation-to-source-code traceability links using latent semantic indexing
Andrian Marcus, Jonathan I. Maletic
May 2003 **Proceedings of the 25th International Conference on Software Engineering**
Publisher: IEEE Computer Society
Full text available: pdf(1.15 MB) Additional Information: full citation, abstract, references, citings, index terms
Publisher Site

An information retrieval technique, latent semantic indexing, is used to automatically identify traceability links from system documentation to program source code. The results of two experiments to identify links in existing software systems (i.e., the LEDA library, and Albergate) are presented. These results are compared with other similar type experimental results of traceability link identification using different types of information retrieval techniques. The method presented proves to give ...

37 Automatic text categorization in terms of genre and author
Efstathios Stamatatos, George Kokkinakis, Nikos Fakotakis
December 2000 **Computational Linguistics**, Volume 26 Issue 4
Publisher: MIT Press
Full text available: pdf(1.30 MB) Additional Information: full citation, abstract, references, citings
Publisher Site

The two main factors that characterize a text are its content and its style, and both can be used as a means of categorization. In this paper we present an approach to text categorization in terms of genre and author for Modern Greek. In contrast to previous stylometric approaches, we attempt to take full advantage of existing natural language processing (NLP) tools. To this end, we propose a set of style markers including analysis-level measures that represent the way in which the input text ha ...

38 Text representation: Word sense disambiguation in information retrieval revisited
Christopher Stokoe, Michael P. Oakes, John Tait
July 2003 **Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval**
Publisher: ACM Press
Full text available: pdf(182.34 KB) Additional Information: full citation, abstract, references, citings, index terms

Word sense ambiguity is recognized as having a detrimental effect on the precision of information retrieval systems in general and web search systems in particular, due to the sparse nature of the queries involved. Despite continued research into the application of automated word sense disambiguation, the question remains as to whether less than 90% accurate automated word sense disambiguation can lead to improvements in retrieval effectiveness. In this study we explore the development and subse ...

Keywords: information retrieval, performance evaluation, word sense disambiguation

39 Fast and effective text mining using linear-time document clustering

Bjornar Larsen, Chinatsu Aone

August 1999 **Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining**

Publisher: ACM Press

Full text available: pdf(897.82 KB) Additional Information: full citation, references, citings, index terms

Keywords: clustering, multi-document summarization, text mining

40 Summarization: The use of unlabeled data to improve supervised learning for text summarization

Massih-Reza Amini, Patrick Gallinari

August 2002 **Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(276.81 KB) Additional Information: full citation, abstract, references, index terms

With the huge amount of information available electronically, there is an increasing demand for automatic text summarization systems. The use of machine learning techniques for this task allows one to adapt summaries to the user needs and to the corpus characteristics. These desirable properties have motivated an increasing amount of work in this field over the last few years. Most approaches attempt to generate summaries by extracting sentence segments and adopt the supervised learning paradigm ...

Keywords: machine learning, semi-supervised learning, text summarization, text-span extraction

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2006 ACM, Inc.
Terms of Usage Privacy Policy Code of Ethics Contact Us

Useful downloads: Adobe Acrobat QuickTime Windows Media Player Real Player

PORTAL USPTO
Subscribe (Full Service) Register (Limited Service, Free) Login
Search: The ACM Digital Library The Guide
sentences and term vector frequency
THE ACM DIGITAL LIBRARY
Feedback Report a problem Satisfaction survey

Terms used **sentences** and **term vector frequency** Found **44,965** of **169,166**

Sort results by relevance
Display results expanded form
Save results to a Binder
Search Tips
Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 141 - 160 of 200 Result page: previous 1 2 3 4 5 6 7 **8** 9 10 next
Best 200 shown Relevance scale

141 Models of translational equivalence among words
I. Dan Melamed
June 2000 **Computational Linguistics**, Volume 26 Issue 2
Publisher: MIT Press
Full text available: pdf(1.89 MB) Additional Information: full citation, abstract, references, citings
Publisher Site

Parallel texts (bitexts) have properties that distinguish them from other kinds of parallel data. First, most words translate to only one other word. Second, bitext correspondence is typically only partial---many words in each text have no clear equivalent in the other text. This article presents methods for biasing statistical translation models to reflect these properties. Evaluation with respect to independent human judgments has confirmed that translation models biased in this fashion are si ...

142 Selective sampling for example-based word sense disambiguation
Atsushi Fujii, Takenobu Tokunaga, Kentaro Inui, Hozumi Tanaka
December 1998 **Computational Linguistics**, Volume 24 Issue 4
Publisher: MIT Press
Full text available: pdf(1.74 MB) Additional Information: full citation, abstract, references
Publisher Site

This paper proposes an efficient example sampling method for example-based word sense disambiguation systems. To construct a database of practical size, a considerable overhead for manual sense disambiguation (overhead for supervision) is required. In addition, the time complexity of searching a large-sized database poses a considerable problem (overhead for search). To counter these problems, our method selectively samples a smaller-sized effective subset from a given example set for use in wor ...

143 Integrating automatic genre analysis into digital libraries
Andreas Rauber, Alexander Müller-Kögler
January 2001 **Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries**
Publisher: ACM Press
Full text available: pdf(672.98 KB) Additional Information: full citation, abstract, references, citings, index terms

With the number and types of documents in digital library systems incr easing, tools for automatically organizing and presenting the content have to be found. While many approaches focus on topic-based organization and structuring, hardly any system incorporates automatic structural analysis and representation. Yet, genre information (unconsciously) forms one of the most distinguishing features in conventional libraries and in information searches. In this paper we present an approach to au ...

Keywords: SOMLib, document clustering, genre analysis, metaphor graphics, self-organizing map (SOM), visualization

144 Correction of errors in a verb modality corpus for machine translation with a machine-learning method

Masaki Murata, Masao Utiyama, Kiyotaka Uchimoto, Hitoshi Isahara, Qing Ma
March 2005 **ACM Transactions on Asian Language Information Processing (TALIP)**, Volume 4 Issue 1
Publisher: ACM Press
Full text available: pdf(342.42 KB) Additional Information: full citation, abstract, references, index terms

In recent years, various types of tagged corpora have been constructed and much research using tagged corpora has been done. However, tagged corpora contain errors, which impedes the progress of research. Therefore, the correction of errors in corpora is an important research issue. In this study we investigate the correction of such errors, which we call *corpus correction.* Using machine-learning methods, we applied corpus correction to a verb modality corpus for machine translation. We u ...

Keywords: corpus correction, machine learning, machine translation, modality corpus

145 Implementing ranking strategies using text signatures

W. Bruce Croft, Pasquale Savino
January 1988 **ACM Transactions on Information Systems (TOIS)**, Volume 6 Issue 1
Publisher: ACM Press
Full text available: pdf(1.59 MB) Additional Information: full citation, abstract, references, citings, index terms, review

Signature files provide an efficient access method for text in documents, but retrieval is usually limited to finding documents that contain a specified Boolean pattern of words. Effective retrieval requires that documents with similar meanings be found through a process of plausible inference. The simplest way of implementing this retrieval process is to rank documents in order of their probability of relevance. In this paper techniques are described for implementing probabilistic ranking ...

146 Web page classification based on k-nearest neighbor approach

Oh-Woog Kwon, Jong-Hyeok Lee
November 2000 **Proceedings of the fifth international workshop on on Information retrieval with Asian languages**
Publisher: ACM Press
Full text available: pdf(653.68 KB) Additional Information: full citation, abstract, references

Automatic categorization is the only viable method to deal with the scaling problem of the World Wide Web. In this paper, we propose a Web page classifier based on an adaptation of k-Nearest Neighbor (k-NN) approach. To improve the performance of k-NN approach, we supplement k-NN approach with a feature selection method and a term-weighting scheme using markup tags, and reform document-document similarity measure used in vector space model. In our experiments on a Korean commercial Web direct ...

Keywords: Web page classification, feature selection, k-nearest neighbor approach, similarity measure, term weighting scheme, text categorization

147 Formal models-1: Discriminative models for information retrieval

Ramesh Nallapati
July 2004 **Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR '04**
Publisher: ACM Press

Full text available: pdf(112.10 KB) Additional Information: full citation, abstract, references, index terms

Discriminative models have been preferred over generative models in many machine learning problems in the recent past owing to some of their attractive theoretical properties. In this paper, we explore the applicability of discriminative classifiers for IR. We have compared the performance of two popular discriminative models, namely the maximum entropy model and support vector machines with that of language modeling, the state-of-the-art generative model for IR. Our experiments on ad-hoc retrie ...

Keywords: discriminative models, machine learning, maximum entropy, pattern classification, support vector machines

148 Information Retrieval: Query-free news search

Monika Henzinger, Bay-Wei Chang, Brian Milch, Sergey Brin

May 2003 **Proceedings of the 12th international conference on World Wide Web**

Publisher: ACM Press

Full text available: pdf(126.70 KB) Additional Information: full citation, abstract, references, citings, index terms

Many daily activities present information in the form of a stream of text, and often people can benefit from additional information on the topic discussed. TV broadcast news can be treated as one such stream of text; in this paper we discuss finding news articles on the web that are relevant to news currently being broadcast.We evaluated a variety of algorithms for this problem, looking at the impact of inverse document frequency, stemming, compounds, history, and query length on the relevance a ...

Keywords: query-free search, web information retrieval

149 Manipulation of trees in information retrieval

Gerard Salton

February 1962 **Communications of the ACM**, Volume 5 Issue 2

Publisher: ACM Press

Full text available: pdf(1.26 MB) Additional Information: full citation, references, citings

150 Recent Studies in Automatic Text Analysis and Document Retrieval

G. Salton

April 1973 **Journal of the ACM (JACM)**, Volume 20 Issue 2

Publisher: ACM Press

Full text available: pdf(1.56 MB) Additional Information: full citation, abstract, references, citings, index terms

Many experts in mechanized text processing now agree that useful automatic language analysis procedures are largely unavailable and that the existing linguistic methodologies generally produce disappointing results. An attempt is made in the present study to identify those automatic procedures which appear most effective as a replacement for the missing language analysis. A series of computer experiments is described, designed to simulate a conventional document retrieval environ ...

151 Taxonomy learning: factoring the structure of a taxonomy into a semantic classification decision

Viktor Pekar, Steffen Staab

August 2002 **Proceedings of the 19th international conference on Computational linguistics - Volume 1**

Publisher: Association for Computational Linguistics

Full text available: pdf(95.55 KB) Additional Information: full citation, abstract, references

The paper examines different possibilities to take advantage of the taxonomic

organization of a thesaurus to improve the accuracy of classifying new words into its classes. The results of the study demonstrate that taxonomic similarity between nearest neighbors, in addition to their distributional similarity to the new word, may be useful evidence on which classification decision can be based.

152 Automatic identification of word translations from unrelated English and German corpora

Reinhard Rapp

June 1999 **Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics**

Publisher: Association for Computational Linguistics

Full text available: pdf(725.23 KB) Additional Information: full citation, abstract, references, citings

Algorithms for the alignment of words in translated texts are well established. However, only recently new approaches have been proposed to identify word translations from non-parallel or even unrelated texts. This task is more difficult, because most statistical clues useful in the processing of parallel texts cannot be applied to non-parallel texts. Whereas for parallel texts in some studies up to 99% of the word alignments have been shown to be correct, the accuracy for non-parallel texts ...

153 Knowledge and representation: Leveraging a common representation for personalized search and summarization in a medical digital library

Kathleen R. McKeown, Noemie Elhadad, Vasileios Hatzivassiloglou

May 2003 **Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries**

Publisher: IEEE Computer Society

Full text available: pdf(116.18 KB) Additional Information: full citation, abstract, references, citings, index terms

Despite the large amount of online medical literature, it can be difficult for clinicians to find relevant information at the point of patient care. In this paper, we present techniques to personalize the results of search, making use of the online patient record as a sophisticated, pre-existing user model. Our work in *PERSIVAL*, a medical digital library, includes methods for re-ranking the results of search to prioritize those that better match the patient record. It also generates summa ...

154 Discovering unexpected information from your competitors' web sites

Bing Liu, Yiming Ma, Philip S. Yu

August 2001 **Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining**

Publisher: ACM Press

Full text available: pdf(1.07 MB) Additional Information: full citation, abstract, references, citings, index terms

Ever since the beginning of the Web, finding useful information from the Web has been an important problem. Existing approaches include keyword-based search, wrapper-based information extraction, Web query and user preferences. These approaches essentially find information that matches the user's explicit specifications. This paper argues that this is insufficient. There is another type of information that is also of great interest, i.e., unexpected information, which is unanticipated by the use ...

Keywords: Information interestingness, Web comparison, Web mining

155 A performance evaluation of similarity measures, document term weighting schemes and representations in a Boolean environment

Terry Noreault, Michael McGill, Matthew B. Koll

June 1980 **Proceedings of the 3rd annual ACM conference on Research and development in information retrieval**

Publisher: Butterworth & Co.

Full text available: pdf(581.27 KB) Additional Information: full citation, references, citings

156 An interpretation of index term weighting schemes based on document components

K. L. Kwok

September 1986 **Proceedings of the 9th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(761.65 KB) Additional Information: full citation, abstract, references, citings

A theory of indexing is presented and is based on viewing a document as constituted of components. A component may be chosen as any run of text unit that can be: (a) judged as to its relevancy property; and (b) considered as independent within the document. By looking at the constituent components of a document in relation to the universe of all components from the collection, we have been able to apply Bayes' decision theory to derive the index term representation for the document, as well ...

157 Applying probabilistic term weighting to OCR text in the case of a large alphabetic library catalogue

Elke Mittendorf, Peter Schäuble, Páraic Sheridan

July 1995 **Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(817.59 KB) Additional Information: full citation, references, citings, index terms

158 Industrial and practical experience track paper session 1: Ranking definitions with supervised learning methods

Jun Xu, Yunbo Cao, Hang Li, Min Zhao

May 2005 **Special interest tracks and posters of the 14th international conference on World Wide Web**

Publisher: ACM Press

Full text available: pdf(328.82 KB) Additional Information: full citation, abstract, references, index terms

This paper is concerned with the problem of definition search. Specifically, given a term, we are to retrieve definitional excerpts of the term and rank the extracted excerpts according to their likelihood of being good definitions. This is in contrast to the traditional approaches of either generating a single combined definition or simply outputting all retrieved definitions. Definition ranking is essential for the task. Methods for performing definition ranking are proposed in this paper, whi ...

Keywords: classification, ordinal regression, search of definitions, text mining, web mining, web search

159 Machine translation vs. dictionary term translation: a comparison for English-Japanese news article alignment

Nigel Collier, Hideki Hirakawa, Akira Kumano

August 1998 **Proceedings of the 17th international conference on Computational linguistics - Volume 1 , Proceedings of the 36th annual meeting on Association for Computational Linguistics - Volume 1**

Publisher: Association for Computational Linguistics , Association for Computational Linguistics

Full text available: pdf(490.59 KB) Publisher Site Additional Information: full citation, abstract, references, citings

Bilingual news article alignment methods based on multi-lingual information retrieval have been shown to be successful for the automatic production of so-called noisy-parallel corpora. In this paper we compare the use of machine translation (MT) to the commonly used dictionary term lookup (DTL) method for Reuter news article alignment in English and Japanese. The results show the trade-off between improved lexical disambiguation

provided by machine translation and extended synonym choice provide ...

160 Document and passage retrieval based on hidden Markov models

Elke Mittendorf, Peter Schäuble

August 1994 **Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: Springer-Verlag New York, Inc.

Full text available: pdf(827.36 KB) Additional Information: full citation, references, citings, index terms

Results 141 - 160 of 200 Result page: previous 1 2 3 4 5 6 7 **8** 9 10 next

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2006 ACM, Inc.
Terms of Usage Privacy Policy Code of Ethics Contact Us

Useful downloads: Adobe Acrobat QuickTime Windows Media Player Real Player